From Single-Race Reporting to Multiple-Race Reporting:
Using Imputation Methods to Bridge the Transition

Nathaniel Schenker, Ph.D.[1] and Jennifer D. Parker, Ph.D.[2]

[1] Corresponding and presenting author. Office of Research and Methodology, National Center for Health Statistics, 6525 Belcrest Road, Hyattsville, MD 20782. Phone: (301) 458-4483. Fax: (301) 458-4031. E-mail: nschenker@cdc.gov.

[2] Infant and Child Health Studies Branch, National Center for Health Statistics, 6525 Belcrest Road, Hyattsville, MD 20782. Phone: (301) 458-4419. Fax: (301) 458-4037. E-mail: jdparker@cdc.gov.

Objective: In 1997, the Office of Management and Budget issued a revised standard for the collection of race information within the federal statistical system. One revision in this standard allows individuals to choose one or more race groups when responding to federal surveys and other federal data collections. The objective of this research is to explore methods that impute single-race categories for those who have given multiple-race responses. Such imputations would be useful when it is desired to conduct analyses involving only single-race categories, such as when changes by race group are being examined over time so that data collected under the old and new standards are being combined.

Methods: The National Health Interview survey has allowed multiple-race responses for several years, while also asking respondents to specify one race as their "main" race. Using National Health Interview Survey data from 1993-1995, we fit logit models that predict main race for respondents in a given multiple-race group. Predictors include age, income, education, Hispanic origin, geography, and contextual variables.

Results: Preliminary results comparing the predicted main race with the reported main race for 1993-1995 suggest that imputation methods that use covariates can have substantially decreased misclassification rates compared to initial methods proposed by the Office of Management and Budget that do not consider covariates, especially for demographic subgroups. Work in progress is examining whether such benefits can be accrued when only contextual covariates are used. Models that only use contextual variables would have the advantage of greater applicability to other data sets, which might not contain useful individual-level covariates. Further research will examine how well models based on the 1993-1995 data predict main race in the National Health Interview Survey data for 1997-1998, as well as how models based on the 1993-1995 data compare to models based on the 1997-1998 data.

Conclusions: Imputation methods have the potential to be very useful in bridging the transition from single-race reporting to multiple-race reporting. In fact, imputation methods have been used previously in a similar problem involving changes in industry and occupation classification schemes between the 1970 and 1980 censuses (Schenker, N., Treiman, D.J., and Weidman, L. (1993), "Analyses of Public-Use Decennial Census Data with Multiply-Imputed Industry and Occupation Codes," *Applied Statistics*, 42, 545-556). To properly reflect the uncertainty due to imputing an unknown single-race category, a technique such as multiple imputation should be used.